Bayesian anomaly detection

Samuel Alan Kossoff Leeney, Will Handley, Dominic Anstey, Eloy de Lera Acedo

14 November 2023

Breakthrough Listen Technosignature Workshop 2023

UNIVERSITY OF CAMBRIDGE

Science and Technology Facilities Council

European Research Council
Established by the European Commission

St Edmund's College
University of Cambridge

Bayesian statistics

Bayes Theorem

"Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail."

Bayes Theorem

"Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail."

- Is Steve a farmer or a librarian?

Bayes Theorem

- Is Steve a farmer or a librarian?
- 90% of people asked guess librarian [Tversky and Kahneman, 1986]
- But there are 20x as many farmers than librarians

Bayes Theorem

We can approach this problem using Bayes theorem.

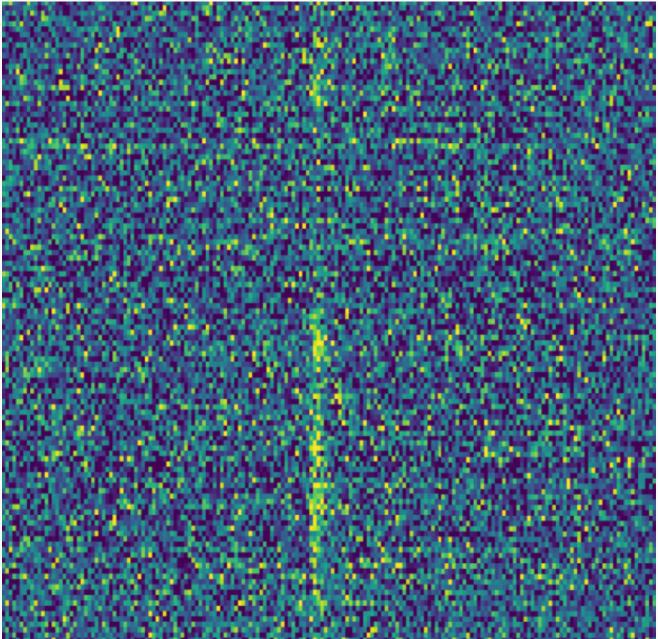
$$\text{likelihood} \times \text{prior} = \text{posterior} \times \text{evidence} \tag{1}$$

$$P(\mathcal{D}|\theta) \times P(\theta) = P(\theta|\mathcal{D}) \times P(\mathcal{D}), \tag{2}$$

$$\mathcal{L} \times \pi = \mathcal{P} \times \mathcal{Z}, \tag{3}$$

Anomaly detection

What do(nt) we look for?



Defining an anomaly sensitive likelihood

a) Generate new likelihood:

$$P(\mathcal{D}_i|\theta) = \begin{cases} \mathcal{L}_i(\theta) & : \text{expected} \\ \Delta^{-1}[0 < \mathcal{D}_i < \Delta] & : \text{anomalous}, \end{cases} \tag{4}$$

b) Ascribe Bernoulli prior:

$$P(\varepsilon_i) = p_i^{(1-\varepsilon_i)}(1-p_i)^{\varepsilon_i}. \tag{5}$$

c) Marginalise over epsilon:

$$P(\mathcal{D}|\theta) = \sum_{\varepsilon \in \{0,1\}^N} P(\mathcal{D}, \varepsilon|\theta) \tag{6}$$

d) Approximate correct mask is most likely

$$P(\mathcal{D}|\theta, \varepsilon_{\max}) \gg \max_j P(\mathcal{D}|\theta, \varepsilon^{(j)}), \tag{7}$$

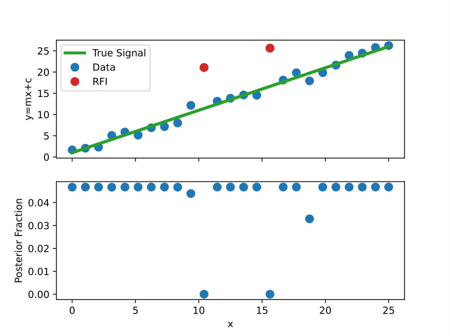
d) Loglikelihood:

$$\log P(\mathcal{D}|\theta) = \sum_i [\log \mathcal{L}_i + \log(1-p_i)]\varepsilon_i^{\max} + [\log p_i - \log \Delta](1-\varepsilon_i^{\max}) \tag{8}$$

Visualising on a simple toy model

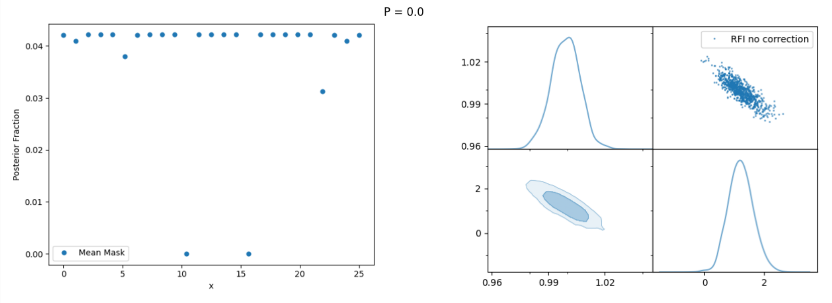
Using numerical sampling techniques to make predictions with $\mathcal{L}$.

- This computes the fraction of posterior believed to 'fit' model given the data.



Probability thresholding condition p

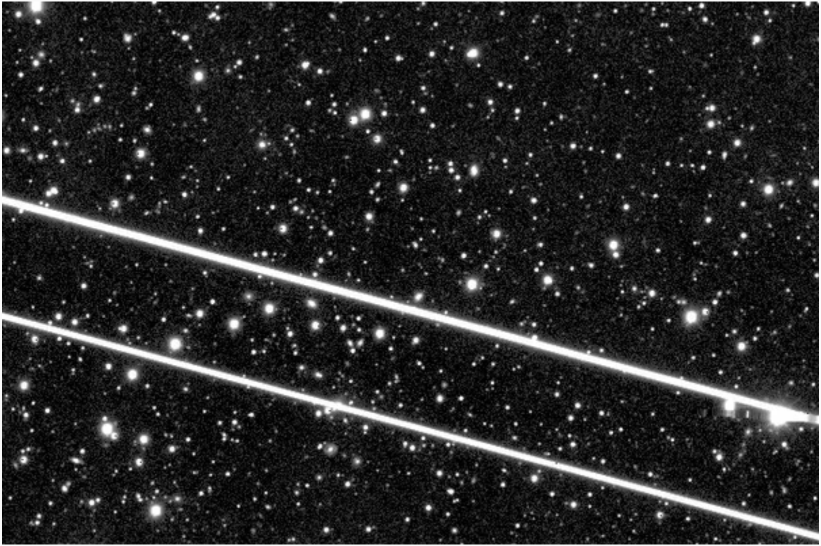
$$\log P(\mathcal{D}|\theta) = \sum_i [\log \mathcal{L}_i + \log(1 - p_i)]\varepsilon^{\max} + [\log p_i - \log \Delta](1 - \varepsilon_i^{\max}), \tag{9}$$



Fully automated anomaly detection

- Putting a prior on p, we can fit it dynamically as a free parameter.
- This fully automates the anomaly detection process.

Initially developed to mitigate for anomalies



Implement with 2 lines of code

```python
def likelihood(theta):
    sig = theta[0]
    logL = -(f_noise - window)**2/sig**2/2 - np.log(2*np.pi*sig**2)/2
    return logL, []
```

```python
def likelihood(theta):
    sig = theta[0]
    logL = -(f_noise - window)**2/sig**2/2 - np.log(2*np.pi*sig**2)/2 + np.log(1-p)
    emax = logL > logp - np.log(delta)
    logPmax = np.where(emax, logL, logp - np.log(delta)).sum()
    return logPmax, []
```

Tutorial @ github.com/samleeney

Read the paper!

PHYSICAL REVIEW D **108,** 062006 (2023)

Bayesian approach to radio frequency interference mitigation

S. A. K. Leeney,[*,§] W. J. Handley,[†,§] and E. de Lera Acedo[‡,§]
The University of Cambridge, Astrophysics Group, Cavendish Laboratory,
J. J. Thomson Avenue, Cambridge, CB3 0HE, United Kingdom

(Received 5 May 2023; accepted 29 August 2023; published 29 September 2023)

Interfering signals such as radio frequency interference from ubiquitous satellite constellations are becoming an endemic problem in fields involving physical observations of the electromagnetic spectrum. To address this we propose a novel data cleaning methodology. Contamination is simultaneously flagged and managed at the likelihood level. It is modeled in a Bayesian fashion through a piecewise likelihood that is constrained by a Bernoulli prior distribution. The techniques described in this paper can be implemented with just a few lines of code.

DOI: 10.1103/PhysRevD.108.062006

arxiv: 2211.15448

Time sensitive anomaly detection

Time sensitive likelihood

- Likelihood from before is extended into two dimensions, becoming

$$\log \mathcal{L}(\theta) = \sum_{ij} \left[\log \mathcal{L}_{ij}(\theta) + \log(1 - p_{ij})\right] \epsilon_{ij} + \left[\log p_{ij} - \log \Delta\right](1 - \epsilon_{ij}) \quad (10)$$

Time sensitive likelihood

- Computation time of the likelihood grows linearly with the number of time bins used in the data set.

$$\log \mathcal{L}(\theta) = \sum_{ij} \left[\log \mathcal{L}_{ij}(\theta) + \log(1 - p_{ij})\right] \epsilon_{ij} + \left[\log p_{ij} - \log \Delta\right](1 - \epsilon_{ij}) \quad (11)$$

Speeding up

- Anstey proposes a solution to this problem in [Anstey et al., 2023].
- Common model per time bin is fit jointly.
- Speeds up fit but not sensitive to transients.

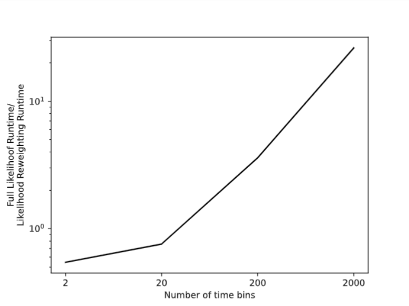
$$
\begin{aligned}
\log \mathcal{L} = & -\frac{1}{2} N_{\mathrm{t}} N_{\nu} \log(2\pi\sigma_{\mathrm{n}}) \\
& - \frac{D}{2{\sigma_{\mathrm{n}}}^2} - \frac{1}{2{\sigma_{\mathrm{n}}}^2} \Bigg[\sum_i \sum_k K_{\mathrm{sq}\,i\,k} F(\nu_i, \theta_{\mathrm{F}\,k})^2 \\
& + \sum_{k_1} \sum_{k_2} \Bigg[\sum_i K_{\mathrm{cross}\,i\,k_1\,k_2} F(\nu_i, \theta_{\mathrm{F}\,k_1}) F(\nu_i, \theta_{\mathrm{F}\,k_2}) \Bigg] \\
& - \mathrm{Tr} \sum_i K_{\mathrm{cross}\,i\,k_1\,k_2} F(\nu_i, \theta_{\mathrm{F}\,k_1}) F(\nu_i, \theta_{\mathrm{F}\,k_2}) \Bigg] \\
& - \frac{1}{{\sigma_{\mathrm{n}}}^2} \sum_i \Bigg[\frac{N_{\mathrm{t}}}{2} T_{\mathrm{S}}(\nu_i, \theta_{\mathrm{S}})^2 - \sum_k T_{\mathrm{D}\cdot\mathrm{K}\,i\,k} F(\nu_i, \theta_{\mathrm{F}\,k}) \\
& - T_{\mathrm{D}\,i} T_{\mathrm{S}}(\nu_i, \theta_{\mathrm{S}}) + \sum_k K_k F(\nu_i, \theta_{\mathrm{F}\,k}) T_{\mathrm{S}}(\nu_i, \theta_{\mathrm{S}}) \Bigg], \qquad (7)
\end{aligned}
$$

Likelihood reweighting

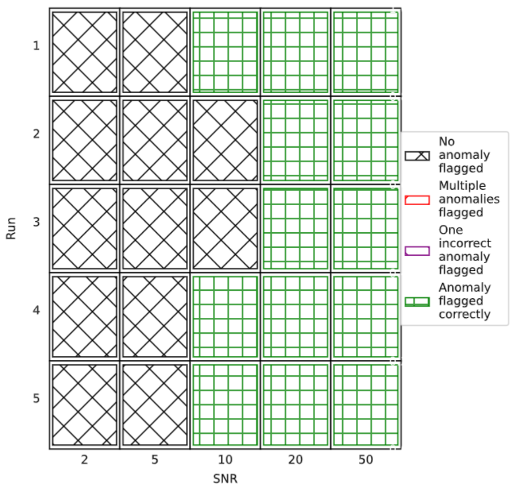
- Introduced in context of gravitational waves by [Payne et al., 2019] [Romero-Shaw et al., 2019].
- Bayesian sampling techniques spend lots of time at tails of distribution.
- Reweighting is essentially a course then fine search, using a simple (fast) then complex (slow) likelihood.

How does this help us?

- We can use the fast method for a coarse scan.
- Then the slower method for refined scan.
- Increases speed massively for larger problems.



Testing on a simple toy example



Read the paper!

Enhanced Bayesian RFI Mitigation and Transient Flagging Using Likelihood Reweighting

Dominic Anstey,[1,2]★ Sam A. K. Leeney,[1,2]
[1]Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge, CB3 0HE, UK
[2]Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
Contamination by Radio Frequency Interference (RFI) is a ubiquitous challenge for radio astronomy. In particular, transient RFI is difficult to detect and avoid, especially in large data sets with many time bins. In this work, we present a Bayesian methodology for time-dependent, transient anomaly mitigation. In general, the computation time for correcting for transient anomalies in time-separated data sets grows proportionally with the number of time bins. We demonstrate that utilising likelihood reweighting can allow our Bayesian anomaly mitigation method to be performed with a computation time close to independent of the number of time bins. In particular, we identify a factor of 25 improvement in computation time for a test case with 2000 time bins. We also demonstrate how this method enables the flagging threshold to be fit for as a free parameter, fully automating the mitigation process. We find that this threshold fitting also prevents overcorrecting of the data in the case of wide priors. Finally, we investigate the potential of the methodology as a transient detector. We demonstrate that the method is able to reliably flag an individual anomalous data point out of 302,000 provided the SNR > 10.

Key words: methods: data analysis – radio continuum: transients

arxiv: 2310.02146

Conclusions

- Fast scans of old data - hopefully find something new!
- 2 lines of code to implement into existing Bayesian systems
- Colaborate with us!

Scan the QR code for the slides or to contact me!

References

Anstey, D., de Lera Acedo, E., and Handley, W. (2023).
Use of time dependent data in bayesian global 21-cm foreground and signal modelling.
Monthly Notices of the Royal Astronomical Society, 520(1):850–865.

Payne, E., Talbot, C., and Thrane, E. (2019).
Higher order gravitational-wave modes with likelihood reweighting.
Physical Review D, 100(12):123017.

Romero-Shaw, I. M., Lasky, P. D., and Thrane, E. (2019).
Searching for eccentricity: signatures of dynamical formation in the first gravitational-wave transient catalogue of ligo and virgo.
Monthly Notices of the Royal Astronomical Society,